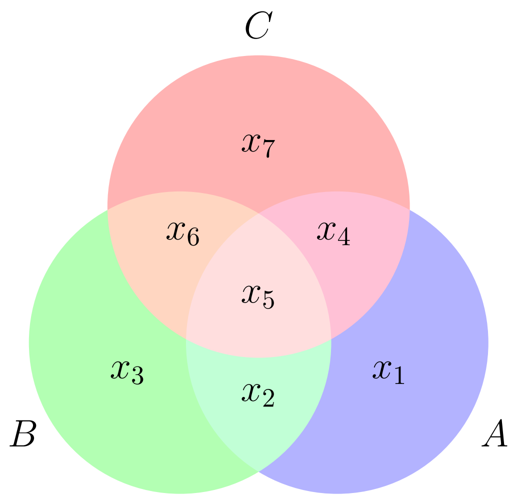
Answer sheet 1

Assignment 1. Define $x_1 = P(A \cap (B \cup C)^c)$, $x_2 = P(A \cap B \cap C^c)$, $x_3 = P(B \cap (A \cup C)^c)$, $x_4 = P(A \cap C \cap B^c)$, $x_5 = P(A \cap B \cap C)$, $x_6 = P(B \cap C \cap A^c)$, $x_7 = P(C \cap (A \cup B)^c)$.



Then, using the information given, we have

$$\begin{aligned}
P(A) &= x_1 + x_2 + x_4 + x_5 = 0.4 && (1)\\
P(B) &= x_2 + x_3 + x_5 + x_6 = 0.7 && (2)\\
P(A \cup B) &= x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 0.8 && (3)\\
P(C \cap (A \cup B)) &= x_4 + x_5 + x_6 = 0.2 && (4)\\
P(B \cap (A \cup C)) &= x_2 + x_5 + x_6 = 0.4 && (5)\\
P(C) &= x_4 + x_5 + x_6 + x_7 = 0.3 && (6)\\
P(B \cap C) &= x_5 + x_6 = 0.2 && (7)
\end{aligned}$$

Using (1)-(3), we get

$$P(A \cap B) = x_2 + x_5 = 0.3 \qquad (8)$$

Using (8) and (5), we get $x_6 = 0.1$. Using (4) and (7), we get $x_4 = 0$. Putting these in (4), we get $x_5 = 0.1$. So, (6) yields $x_7 = 0.1$, and (5) yields $x_2 = 0.2$. From (1) and (2), we now get $x_1 = 0.1$ and $x_3 = 0.3$.

(a) The probability that exactly two of A, B and C occur equals $x_2 + x_4 + x_6 = 0.3$.

(b) The probability that none of A, B and C occur equals $1 - \sum_{k=1}^{7} x_k = 0.1$.

(c) The probability that A and exactly one of B and C occur equals $x_2 + x_4 = 0.2$.

Assignment 2. Denote by 'H' and 'T' the events that a head and a tail occurs, respectively. A sample point $\{x, y, z\}$ with $x, y, z \in \{H, T\}$ will denote that x occurs for A, y occurs for B and z occurs for C.

(a) $\Omega = \{\{H,H,H\}, \{H,H,T\}, \{H,T,H\}, \{H,T,T\}, \{T,H,H\}, \{T,H,T\}, \{T,T,H\}, \{T,T,T\}\}$.

(b) Assuming the each coin is fair, all the sample points may be assigned equal probabilities, i.e., $P(\{x, y, z\}) = 1/8$ for all $\{x, y, z\} \in \Omega$.

(c) P(A wins on first toss) $=$ P($\{H,T,T\} \cup \{T,H,H\}$) $= 1/4$. The probabilities are the same

for B as well as C winning on the first toss.
(d) P(no winner on first toss) = $\mathrm{P}(\{H,H,H\} \cup \{T,T,T\}) = 1/4$.
(e) Let F and G denote the events that A wins on first toss and that the winner is decided on first toss, respectively. Then, $\mathrm{P}(G) = \mathrm{P}(A$ wins on first toss) + $\mathrm{P}(B$ wins on first toss) + $\mathrm{P}(C$ wins on first toss) $= 3/4$. Since $F \subset G$, we have $\mathrm{P}(F \cap G) = \mathrm{P}(F) = 1/4$. So, $\mathrm{P}(F \mid G)$ $= \mathrm{P}(F \cap G)/\mathrm{P}(G) = 1/3$.
(Intuition : since all events are equally likely, if there is a winner on the first toss, then it has to be one of A, B and C with equal probability. This probability is obviously $1/3$*).*

Assignment 3. (a) If the contestant doesn't alter his choice, his chances of winning are the same as before : 1/3.
(b) The contestant may be completely confused and decide to ignore the information of the second door open. He neither sticks to, nor alter his original choice, he rather picks at random one of the two yet unopened doors (e.g. by a coin flip). Then his chance of winning is 1/2.
(c) If the contestant alters his choice to door C, his chances increase to 2/3. Note that among all the possible scenarios, the only scenario leading to him loosing, when the choice is altered, is the scenario in which his original choice was right. Probability that his original choice was right is of course 1/3.
In cases (a) and (c), the only source of randomness is the contestant's original choice. The probability space is a triplet $(\Omega, \mathcal{F}, \mathbb{P})$ (sample space, σ-algebra of events, probability measure), where

$$\begin{aligned}\Omega = \{&[\text{the contestant chose the first door}],\\ &[\text{the contestant chose the second door}],\\ &[\text{the contestant chose the third door}]\},\end{aligned}$$

$\mathcal{F} = 2^\Omega$ and probability $\mathbb{P}$ is given by $\mathbb{P}(A) = \frac{1}{3}$ for any elementary event ω, i.e. any element of Ω. The random variable X is a function that maps the event [the contestant chose the i-th door] to number i, for $i = 1, 2, 3$. It's very natural to forget about probability space and concentrate rather on X and its distribution given by $\mathbb{P}(X = i) = 1/3$ for $i = 1, 2, 3$, which is a discrete uniform distribution on the set $\{1, 2, 3\}$. It should be clear it doesn't matter whether the contender alters his choice or not, probability space remains the same, the only thing that changes is, what choice(s) is (are) the winning one(s).
In case (b), the probability of winning is independent of the original choice. The sample space is reduced by one event (the opened door are now out of the picture), and everything depends on the coin flip.
Note : This is the famous Monty Hall problem. In our solution, we considered e.g. the position of the car deterministic, and the only source of randomness was the contestant's choice. We could also consider the position of the car random and the choice deterministic. Or we could consider both the choice and the position random, independent of each other, and even add another source of randomness – which door does the host open. The sample space would then be a product space of 3 x 3 x 3 events, some of them having probability zero, for example. This is done on Wikipedia. In all cases, results are the same, and the probabilistic models justifiable.

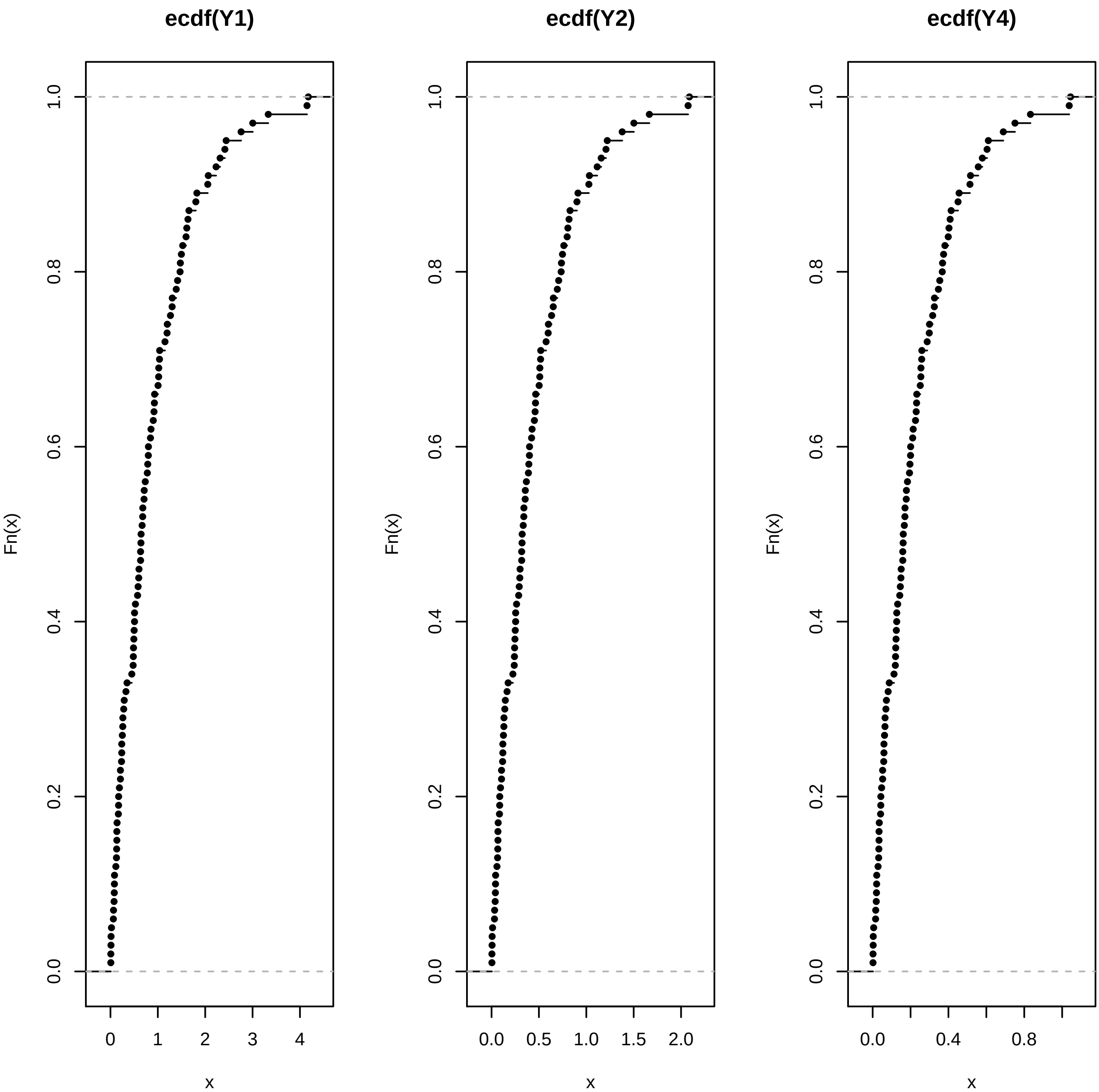
Assignment 4. Denote by Pos and Neg the events that the test result is positive and negative, respectively. Also, denote by D and ND the events that the person actually has the disease and does not have it, respectively. We need to find $\mathrm{P}(D \mid Pos)$. It is given that

$\mathrm{P}(Pos \mid ND) = 1/100$, $\mathrm{P}(Neg \mid D) = 2/100$ and $\mathrm{P}(D) = 1/1000$.
Now, $\mathrm{P}(Pos) = \mathrm{P}(Pos \cap D) + \mathrm{P}(Pos \cap ND)$
$= \mathrm{P}(D)$ - $\mathrm{P}(Neg \cap D) + \mathrm{P}(Pos \cap ND)$
$= \mathrm{P}(D)$ - $\mathrm{P}(D) \times \mathrm{P}(Neg \mid D) + \mathrm{P}(ND) \times \mathrm{P}(Pos \mid ND)$
$= (1/1000) - (1/1000) \times (2/100) + (999/1000) \times (1/100) = 1097/(1000 \times 100) \approx 11/1000$.
The above calculations show that $\mathrm{P}(Pos \cap D) = (1/1000) - (1/1000) \times (2/100) = 98/(1000 \times 100) \approx 1/1000$.
So, we have $\mathrm{P}(D \mid Pos) = \mathrm{P}(Pos \cap D)/\mathrm{P}(Pos) = 98/1097 \approx 1/11$.
Further, $\mathrm{P}(ND \mid Neg) = \mathrm{P}(Neg \cap ND)/\mathrm{P}(Neg)$
$= [\mathrm{P}(ND)$ - $\mathrm{P}(Pos \cap ND)]/[1$ - $\mathrm{P}(Pos)]$
$= [(999/1000) - (999/1000) \times (1/100)]/[1 - 1097/(1000 \times 100)]$
$= 98901/98903 \approx 1$.
Thus, although the chances of false positives (namely, $\mathrm{P}(Pos \mid ND)$) and false negatives (namely, $\mathrm{P}(Neg \mid D)$) from the test procedure are very small, it is not a good idea to base an affirmative diagnosis of the disease on this test alone.

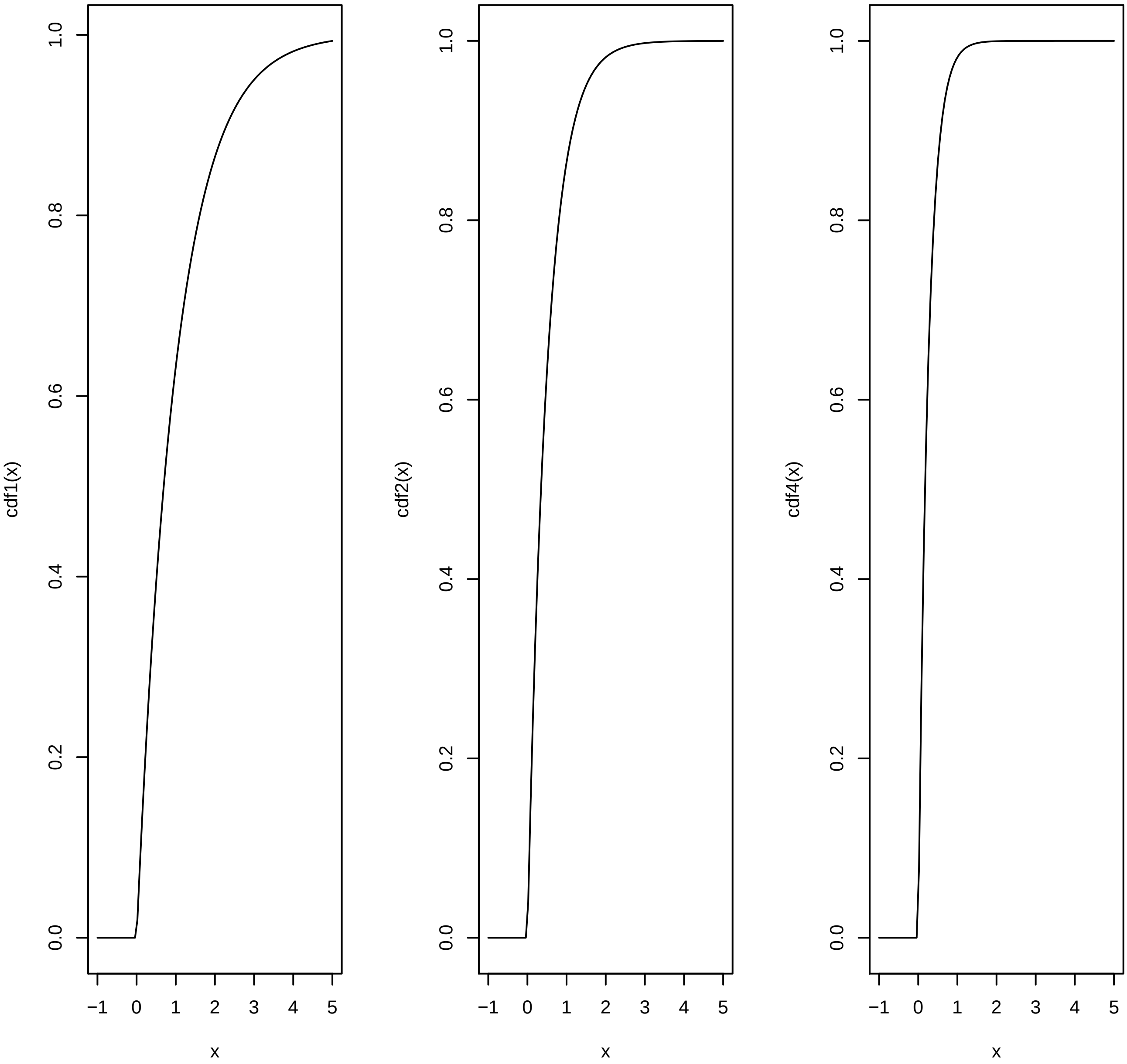
Assignment 5. (e) These plots show the empirical distribution functions

$$t \mapsto \frac{1}{100} \sum_{i=1}^{100} \mathbf{1}\{Y_i \leq t\}$$

for each of the vectors $Y1$, $Y2$ and $Y4$. It is a nondecreasing right-continuous step function (piecewise constant) that equals 0 for t (strictly) smaller than the minimum of the sample, and 1 for t (weakly) larger than the maximum of the sample.



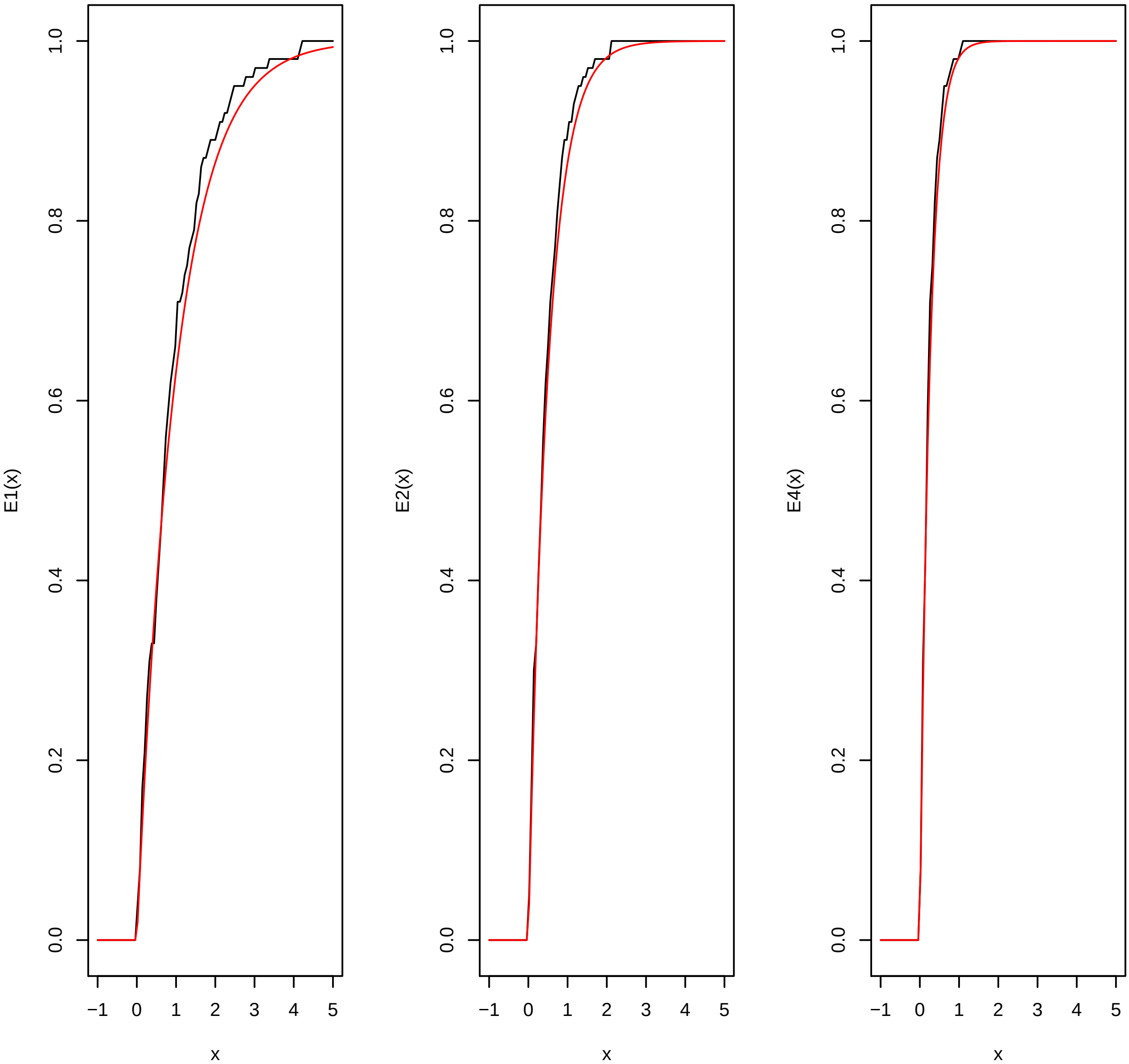
(g) These are plots of the functions F_λ for the parameter values 1, 2, and 4 for λ.



(i) These figures combine the two previous plots. We see that the empirical distribution function for $Y\lambda$ is very close to the function F_λ. The figures imply that $Y\lambda$ follows the distribution F_λ, i.e. Exp(λ). And indeed, since $X \sim$Unif(0,1),

$$P(Y\lambda \le t) = P(-\log(1-X)/\lambda \le t) = P(X \le 1 - \exp(-\lambda t)) = F_\lambda(t).$$

The fact that the empirical distribution function approaches the distribution function itself is known as the *Glivenko–Cantelli* theorem.



(j) q_λ is the inverse of F_λ on $(0, \infty)$.
(k) Yes. This is so because $Y = q(X) = F^{-}(X) = F^{-1}(X)$ is such that

$$P(Y \leq t) = P(X \leq F(t)) = F(t).$$